Seminar on Advanced Topics in Mathematics

Solving Polynomial Equations

5 December 2006

Dr. Tuen Wai Ng, HKU

What do we mean by solving an equation ?

Example 1. Solve the equation $x^2 = 1$.

$$\begin{aligned} x^2 &= 1 \\ x^2 - 1 &= 0 \\ (x-1)(x+1) &= 0 \\ x &= 1 \text{ or } = -1 \end{aligned}$$

- Need to check that in fact $(1)^2 = 1$ and $(-1)^2 = 1$.

Exercise. Solve the equation

$$\sqrt{x} + \sqrt{x-a} = 2$$

where a is a positive real number.

What do we mean by solving a polynomial equation ?

Meaning I:

Solving polynomial equations: *finding numbers* that make the polynomial take on the value zero when they replace the variable.

• We have discovered that x, which is something we didn't know, turns out to be 1 or -1.

Example 2. Solve the equation $x^2 = 5$.

$$\begin{aligned} x^2 &= 5 \\ x^2 - 5 &= 0 \\ (x - \sqrt{5})(x + \sqrt{5}) &= 0 \\ x &= \sqrt{5} \text{ or } -\sqrt{5} \end{aligned}$$

• But what is $\sqrt{5}$? Well, $\sqrt{5}$ is the positive real number that square to 5.

• We have "learned" that the positive solution to the equation $x^2 = 5$ is the positive real number that square to 5 !!!

• So there is a sense of circularity in what we have done here.

• Same thing happens when we say that i is a solution of $x^2 = -1$.

What are “solved” when we solve these equations ?

- The equations $x^2 = 5$ and $x^2 = -1$ draw the attention to an inadequacy in a certain number system (it does not contain a solution to the equation).

- One is therefore driven to extend the number system by introducing, or ‘adjoining’, a solution.

- Sometimes, the extended system has the good algebraic properties of the original one, e.g. addition and multiplication can be defined in a natural way.

- These extended number systems (e.g. $\mathbb{Q}(\sqrt{5})$ or $\mathbb{Q}(i)$) have the added advantage that more equations can be solved.

Consider the equation

$$x^2 = x + 1.$$

• By completing the square, or by applying the formula, we know that the solutions are $\frac{1+\sqrt{5}}{2}$ or $\frac{1-\sqrt{5}}{2}$.

• It is certainly not true by definition that $\frac{1+\sqrt{5}}{2}$ is a solution of the equation.

• What we have done is to take for granted that we can solve the equation $x^2 = 5$ (and similar ones) and to use this interesting ability to solve an equation which is not of such a simple form.

• When we solve the quadratic, what we are actually showing that the problem can be reduced to solving a particularly simple quadratic $x^2 = c$.

What do we mean by solving a polynomial equation ?

Meaning II:

Suppose we can solve the equation $x^n = c$, i.e. taking roots, try to express the the roots of a degree n polynomial using only the usual algebraic operations (addition, subtraction, multiplication, division) and application of taking roots.

• In this sense, one can solve any polynomials of degree 2,3 or 4 and this is in general impossible for polynomials of degree 5 or above.

• The Babylonians (about 2000 B.C.) knew how to solve specific quadratic equations.

• The solution formula for solving the quadratic equations was mentioned in the *Bakshali Manuscript* written in India between 200 BC and 400 AD.

• Based on the work of Scipione del Ferro and Nicolo Tartaglia, Cardano published the solution formula for solving the cubic equations in his book *Ars Magna* (1545).

• Lodovico Ferrari, a student of Cardano discovered the solution formula for the quartic equations in 1540 (published in *Ars Magna* later).

• The formulae for the cubic and quartic are complicated, and the methods to derive them seem ad hoc and not memorable.

Solving polynomial equations using circulant matrices

D. Kalman and J.E. White, Polynomial Equations and Circulant Matrices, The American Mathematical Monthly, **108**, no.9, 821-840, 2001.

Circulant matrices. An $n \times n$ circulant matrix is formed from any n-vector by cyclically permuting the entries. For example, starting with $[a\ b\ c]$ we can generate the 3×3 circulant matrix

$$C = \begin{bmatrix} a & b & c \\ c & a & b \\ b & c & a \end{bmatrix} . \tag{1}$$

- Circulant matrices have constant values on each downward diagonal, that is, along the lines of entries parallel to the main diagonal.

The eigenvalues and eigenvectors of circulant matrices are very easy to compute using the nth roots of unity.

- For the 3×3 matrix C in (1), we need the cube roots of unity:

$$1, \omega = (-1 + i\sqrt{3})/2 \quad \text{and} \quad \omega^2 = \overline{\omega}.$$

- Direct computations show that the eigenvalues of C are $a + b + c$, $a + b\omega + c\omega^2$, and $a + b\overline{\omega} + c\overline{\omega}^2$, with corresponding eigenvectors $(1, 1, 1)^T$, $(1, \omega, \omega^2)^T$, and $(1, \overline{\omega}, \overline{\omega}^2)^T$.

- This result can be generalized to higher dimensions $(n \geq 3)$.

To begin with, we define a distinguished circulant matrix W with first row $(0, 1, 0, \ldots, 0)$. W is just the identity matrix with its top row moved to the bottom, e.g. for $n = 4$,

$$W = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad W^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad W^3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Direct checking shows that

i) Note that $W^T = W^{-1}$ (i.e. W is an $orthogonal$ matrix).

ii) The characteristic polynomial for W is $p(t) = det(tI - W) = t^n - 1$, and hence the eigenvalues of W are the nth roots of unity.

iii) For each nth root of unity λ, $\mathbf{v}_\lambda = (1, \lambda, \lambda^2, \ldots, \lambda^{n-1})$ is an associated eigenvector.

If C is any $n \times n$ circulant matrix, use its first row $[a_0 \ a_1 \ a_2 \ \ldots \ a_{n-1}]$ to define a polynomial $q(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_{n-1} t^{n-1}$.

i) Then $C = q(W) = a_0 I + a_1 W + a_2 W^2 + \cdots + a_{n-1} W^{n-1}$.

For example, when $n = 4$,

$$a_0 I = \begin{bmatrix} a_0 & 0 & 0 & 0 \\ 0 & a_0 & 0 & 0 \\ 0 & 0 & a_0 & 0 \\ 0 & 0 & 0 & a_0 \end{bmatrix}, \quad a_1 W = \begin{bmatrix} 0 & a_1 & 0 & 0 \\ 0 & 0 & a_1 & 0 \\ 0 & 0 & 0 & a_1 \\ a_1 & 0 & 0 & 0 \end{bmatrix}$$

$$a_2 W^2 = \begin{bmatrix} 0 & 0 & a_2 & 0 \\ 0 & 0 & 0 & a_2 \\ a_2 & 0 & 0 & 0 \\ 0 & a_2 & 0 & 0 \end{bmatrix}, \quad a_3 W^3 = \begin{bmatrix} 0 & 0 & 0 & a_3 \\ a_3 & 0 & 0 & 0 \\ 0 & a_3 & 0 & 0 \\ 0 & 0 & a_3 & 0 \end{bmatrix}.$$

Therefore, $q(W) = a_0 I + a_1 W + a_2 W^2 + a_3 W^3$ is equal to

$$C = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 \\ a_3 & a_0 & a_1 & a_2 \\ a_2 & a_3 & a_0 & a_1 \\ a_1 & a_2 & a_3 & a_0 \end{bmatrix}$$

ii) For any nth root of unity $\lambda, q(\lambda)$ is an eigenvalue of $C = q(W)$.

[Indeed, if $W\mathbf{v} = \lambda\mathbf{v}$, then $W^k\mathbf{v} = \lambda^k\mathbf{v}$ and hence $q(W)\mathbf{v} = q(\lambda)\mathbf{v}$.]

Example 3. Consider the circulant matrix

$$C = \begin{bmatrix} 1 & 2 & 1 & 3 \\ 3 & 1 & 2 & 1 \\ 1 & 3 & 1 & 2 \\ 2 & 1 & 3 & 1 \end{bmatrix} .$$

Read the polynomial q from the first row of C:

$$q(t) = 1 + 2t + t^2 + 3t^3.$$

Here, with $n = 4$, the nth roots of unity are ± 1 and $\pm i$.

The eigenvalues of C are now computed as

$$q(1) = 7,\ q(-1) = -3,\ q(i) = -i, \text{ and } \quad q(-i) = i.$$

The corresponding eigenvectors are

$$\begin{aligned} v(1) &= (1, 1, 1, 1), \\ v(-1) &= (1, -1, 1, -1), \\ v(i) &= (1, i, -1, -i), \text{ and} \\ v(-i) &= (1, -i, -1, i). \end{aligned}$$

Can check that the characteristic polynomial of C is

$$\det(tI - C) = p(t) = t^4 - 4t^3 - 20t^2 - 4t - 21.$$

Summary

Start with any circulant matrix C, one can generate both the $roots$ and $coefficients$ of a polynomial p.

Here, the polynomial p is the characteristic polynomial of C; the coefficients can be obtained from the identity $p(t) = \det(tI - C)$; the roots, i.e., the eigenvalues of C, can be found by applying q to the nth roots of unity.

This perspective leads to a unified method for solving general quadratic, cubic, and quartic equations.

In fact, given a polynomial p, we try to find a corresponding circulant C having p as its characteristic polynomial. The first row of C then defines a different polynomial q, and the roots of p are obtained by applying q to the nth roots of unity.

Solving polynomial equations using circulant matrices.

Quadratics. Let's consider a general quadratic polynomial,

$$p(t) = t^2 + \alpha t + \beta.$$

We also consider a general 2×2 circulant

$$C = \begin{bmatrix} a & b \\ b & a \end{bmatrix} \ .$$

The characteristic polynomial of C is

$$\det \begin{bmatrix} t - a & -b \\ -b & t - a \end{bmatrix} = t^2 - 2at + a^2 - b^2.$$

We must find a and b so that this characteristic polynomial equals p, so

$$\begin{array}{rcl} -2a & = & \alpha \\ a^2 - b^2 & = & \beta. \end{array}$$

Solving this system gives $a = -\alpha/2$ and $b = \pm\sqrt{\alpha^2/4 - \beta}$. To proceed, we require only one solution of the system, and for convenience define b with the positive sign, so

$$C = \begin{bmatrix} -\alpha/2 & \sqrt{\alpha^2/4 - \beta} \\ \sqrt{\alpha^2/4 - \beta} & -\alpha/2 \end{bmatrix}$$

and

$$q(t) = \frac{-\alpha}{2} + t\sqrt{\frac{\alpha^2}{4} - \beta}.$$

The roots of the original quadratic are now found by applying q to the two square roots of unity:

$$\begin{aligned} q(1) &= \frac{-\alpha}{2} + \sqrt{\frac{\alpha^2}{4} - \beta} \\ q(-1) &= \frac{-\alpha}{2} - \sqrt{\frac{\alpha^2}{4} - \beta}. \end{aligned}$$

• Observe that defining b with the opposite sign produces the same roots of p, although the values of $q(1)$ and $q(-1)$ are exchanged.

Cubics. A parallel analysis works for cubic polynomials.

We first notice that by simple algebra, for $p(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_1 x + \alpha_0$, the substitution $y = x - \alpha_{n-1}/n$ eliminates the term of degree $n - 1$.

Therefore, we only need to consider cubic polynomials of the form

$$p(t) = t^3 + \beta t + \gamma.$$

For a general 3×3 circulant matrix

$$C = \begin{bmatrix} a & b & c \\ c & a & b \\ b & c & a \end{bmatrix} ,$$

we want to find a, b, and c so that p is the characteristic polynomial of C.

Since sum of roots of p is zero, na which is the sum of eigenvalues of C is also equal to zero.

Therefore,

$$C = \begin{bmatrix} 0 & b & c \\ c & 0 & b \\ b & c & 0 \end{bmatrix} ,$$

and its characteristic polynomial is given by

$$\det \begin{bmatrix} t-0 & -b & -c \\ -c & t-0 & -b \\ -b & -c & t-0 \end{bmatrix} = t^3 - b^3 - c^3 - 3bct.$$

This equals p if

$$\begin{aligned} b^3 + c^3 &= -\gamma \\ 3bc &= -\beta \ . \end{aligned} \tag{4}$$

$$\begin{aligned} b^3 + c^3 &= -\gamma \\ 3bc &= -\beta\ . \end{aligned} \tag{4}$$

To complete the solution of the original equation, we must solve this system for b and c, and then apply $q(x) = bx + cx^2$ to the cube roots of unity.

That is, for any a and b satisfying (4), we obtain the roots of p as $q(1) = b + c, q(\omega) = b\omega + c\omega^2$, and $q(\overline{w}) = b\overline{\omega} + c\overline{w}^2$.

Thinking of the unknowns as b^3 and c^3 makes (4) quite tractable. Indeed, dividing the second equation by 3 and cubing, we get

$$\begin{aligned} b^3 + c^3 &= -\gamma \\ b^3c^3 &= -\tfrac{\beta^3}{27}\ . \end{aligned}$$

Observe that b^3 and c^3 are the roots of the quadratic equation $x^2 + \gamma x - \beta^3/27 = 0$, and so are given by

$$\frac{-\gamma \pm \sqrt{\gamma^2 + 4\beta^3/27}}{2}. \tag{5}$$

At this point, it is tempting to write

$$\begin{aligned} b &= \left[\frac{-\gamma + \sqrt{\gamma^2 + 4\beta^3/27}}{2}\right]^{1/3} \\ c &= \left[\frac{-\gamma - \sqrt{\gamma^2 + 4\beta^3/27}}{2}\right]^{1/3} \quad . \end{aligned} \tag{6}$$

$$\begin{aligned} b &= \left[\frac{-\gamma + \sqrt{\gamma^2 + 4\beta^3/27}}{2}\right]^{1/3} \\ c &= \left[\frac{-\gamma - \sqrt{\gamma^2 + 4\beta^3/27}}{2}\right]^{1/3} \quad . \end{aligned} \tag{6}$$

• This is perfectly valid when all of the operations involve only positive real numbers. In the larger domain of complex numbers there is some ambiguity associated with the extraction of square and cube roots.

In this case, define b by (6), using *any* of the possible values of the necessary square and cube roots, and then take $c = -\beta/(3b)$. That produces a solution to (4), and leads to the roots of p given by

$q(1) = b + c, q(\omega) = b\omega + c\omega^2$, and $q(\overline{w}) = b\overline{\omega} + c\overline{w}^2$.

• All choices for b result in the same roots.

Quartics. We outline the circulant solution of the quartic equation. We only need to consider quartic polynomials of the form

$$p(t) = t^4 + \beta t^2 + \gamma t + \delta,$$

and to avoid a trivial case, we assume that not all of β, γ, and δ vanish.

We seek a circulant matrix

$$C = \begin{bmatrix} 0 & b & c & d \\ d & 0 & b & c \\ c & d & 0 & b \\ b & c & d & 0 \end{bmatrix}$$

with characteristic polynomial equal to p.

The characteristic polynomial of C is

$$\det \begin{bmatrix} t & -b & -c & -d \\ -d & t & -b & -c \\ -c & -d & t & -b \\ -b & -c & -d & t \end{bmatrix} = \begin{array}{l} t^4 - (4bd + 2c^2)t^2 - 4c(b^2 + d^2)t \\ \\ +c^4 - b^4 - d^4 - 4bdc^2 + 2b^2d^2. \end{array}$$

Equating this with $p(t) = t^4 + \beta t^2 + \gamma t + \delta$, produces the system

$$\begin{array}{rcl} 4bd + 2c^2 & = & -\beta \\ 4c(b^2 + d^2) & = & -\gamma \\ c^4 - b^4 - d^4 - 4bdc^2 + 2b^2d^2 & = & \delta\ . \end{array} \tag{7}$$

$$\begin{array}{rcl} 4bd + 2c^2 & = & -\beta \\ 4c(b^2 + d^2) & = & -\gamma \\ c^4 - b^4 - d^4 - 4bdc^2 + 2b^2d^2 & = & \delta\ . \end{array} \tag{7}$$

Now notice that the first and second equations in this system determine bd and $b^2 + d^2$ in terms of c. This inspires us to rewrite the third equation in the form

$$c^4 - (b^2 + d^2)^2 + 4(bd)^2 - 4bdc^2 = \delta$$

and hence to obtain an equation in c alone:

$$c^4 - \frac{\gamma^2}{16c^2} + \frac{(\beta + 2c^2)^2}{4} + (2c^2 + \beta)c^2 = \delta.$$

This simplifies to

$$c^6 + \frac{\beta}{2}c^4 + \left(\frac{\beta^2}{16} - \frac{\delta}{4}\right)c^2 - \frac{\gamma^2}{64} = 0, \tag{8}$$

which is a cubic polynomial equation in c^2, and in principle is solvable by the methods already in hand.

This leads to a nonzero value for c (since β, γ, and δ are not all 0), and it is then straightforward to find corresponding values for b and d so that (7) is satisfied.

In this way we have constructed the circulant matrix

$$C = bW + cW^2 + dW^3 = q(W),$$

whose eigenvalues are the roots of p.

They are computed by applying q to the fourth roots of unity:

$$\begin{aligned} q(1) &= b + c + d \\ q(-1) &= -b + c - d \\ q(i) &= -c + i(b - d) \\ q(-i) &= -c - i(b - d). \end{aligned}$$

This completes the solution of the quartic, and the circulant approach to solving low degree polynomial equations.

- How about polynomials of higher degree ?

We know that a general solution by radicals is not possible for equations beyond the quartic, but why does the circulant method fail ?

- A natural first question is whether every monic polynomial p can be realized as the characteristic polynomial q of a circulant matrix C.

- The answer is yes.

If p is a monic polynomial of degree n with zeros $z_1, ..., z_n$, the question is the same as asking whether one can always find a polynomial q of degree $n-1$ such that for each $1 \leq k \leq n$,

$$q(\omega^k) = z_k,$$

where $\omega^0, \omega^1, \ldots, \omega^{n-1}$ are the nth roots of unity.

If such q exists and is equal to $a_0 + a_1 t + \cdots + a_{n-1} t^{n-1}$, then we can generate C by the first row vector

$$(a_0, a_1, \ldots, a_{n-1}).$$

Since the eigenvalues of C are $q(\omega^k) = z_k$, the characteristic polynomial of C is equal to p.

Existence and uniqueness of q

Let P_{n-1} be the vector space of polynomials of degree at most $n-1$. Define the linear map $L : P_{n-1} \to \mathbb{C}^n$ by

$$L(q) = (q(\omega^0), q(\omega^1), ..., q(\omega^{n-1})).$$

Observe that if $L(q) = (0, 0, ..., 0)$, then $q \in P_{n-1}$ vanishes at the n *distinct* points $\omega^0, \omega^1, \ldots, \omega^{n-1}$, therefore q must be the zero polynomial.

Thus, the kernel of L is zero and hence L is injective. Since both P_{n-1} and $\mathbb{C}^n$ have the same dimension, n, by the dimension formula, the dimension of the image $L(P_{n-1})$ must be equal to n and hence L is surjective.

- L is bijective implies that q exists and is unique.

If p is a monic polynomial of degree n with zeros $z_1, ..., z_n$, we now know that there exists a unique polynomial q of degree $n-1$ such that for each $1 \leq k \leq n$,

$$q(\omega^k) = z_k.$$

Therefore, the circulant method allows us to express the roots of p in terms of the roots of unity and the coefficients of q.

• For $2 \leq n \leq 4$, we have seen that all the coefficients of q can be expressed in terms of radicals of the coefficients of p, therefore all polynomials of degree at most four can be solved by radicals.

• Since there are polynomials with rational coefficients whose roots cannot be expressed in terms of radicals, in general, the circulant matrix entries for a given polynomial may likewise not be expressible in terms of radicals.

The first attempt to unify solutions to quadratic, cubic and quartic equations date at least to Lagrange 's work in "Réflexions sur la résolution algébrique des équations" (1770/1771).

Lagrange's analysis characterized the general solutions of the cubic and quartic cases in terms of permutations of the roots, laying a foundation for the independent demonstrations by Abel and Galois of the impossibility of solutions by radicals for general fifth degree or higher equations.

Abel's Theorem (1824). *The generic algebraic equation of degree higher than four is not solvable by radicals, i.e., formulae do not exist for expressing roots of a generic equation of degree higher than four in terms of its coefficients by means of operations of addition, subtraction, multiplication,division, raising to a natural power, and extraction of a root of natural degree.*

Abel's Proof

Abel's idea was that if some finite sequence of rational operations and root extractions applied to the coefficients produces a root of the equation

$$x^5 - ax^4 + bx^3 - cx^2 + dx - e = 0,$$

the final result must be expressible in the form

$$x = p + R^{\frac{1}{m}} + p_2 R^{\frac{2}{m}} + \cdots + p_{m-1} R^{\frac{m-1}{m}},$$

where $p, p_2, \ldots, p_{m-1}$, and R are also formed by rational operations and root extractions applied to the coefficients, m is a prime number, and $R^{1/m}$ is not expressible as a rational function of the coefficients $a, b, c, d, e, p, p_2, \ldots, p_{m-1}$.

By straightforward reasoning on a system of linear equations for the coefficients p_j, he was able to show that R is a $symmetric$ function of the $roots$, and hence that $R^{1/m}$ must assume exactly m different values as the roots are permuted.

Moreover, since there are $5! = 120$ permutations of the roots and m is a prime, it followed that $m = 2$ or $m = 5$, the case $m = 3$ having been ruled out by Cauchy.

The hypothesis that $m = 5$ led to certain equation in which the left-hand side assumed only five values while the right-hand side assumed 120 values as the roots were permuted.

Then the hypothesis $m = 2$ led to a similar equation in which one side assumed 120 values and the other only 10.

Abel concluded that the hypothesis that there exists an algorithm for solving the equation was incorrect.

P. Pesic, *Abel's proof. An essay on the sources and meaning of mathematical unsolvability*. MIT Press, Cambridge, MA, 2003.

C. Houzel, The work of Niels Henrik Abel. *The legacy of Niels Henrik Abel*, 21–177, Springer, Berlin, 2004

The Abel Prize

The Niels Henrik Abel Memorial Fund was established on 1 January 2002, to award the Abel Prize for outstanding scientific work in the field of mathematics.

The prize amount is 6 million NOK (about 750,000 Euro) and was awarded for the first time on 3 June 2003.

Abel Prize Laureates: Jean-Pierre Serre (2003), Sir Michael Francis Atiyah and Isadore M. Singer (2004), Peter D. Lax (2005), Lennart Carleson (2006).

Solution of the general quintic by elliptic integrals.

• In 1844, Ferdinand Eisenstein showed that the general quintic equation could be solved in terms of a function $\chi(\lambda)$ that satisfies the special quintic equation

$$\big(\chi(\lambda)\big)^5 + \chi(\lambda) = \lambda.$$

This function is in a sense an analog of root extraction, since the square root function φ and the cube root function ψ satisfy the equations

$$\big(\varphi(\lambda)\big)^2 = \lambda,\ \big(\psi(\lambda)\big)^3 = \lambda.$$

• In 1858 Hermite and Kronecker showed (independently) that the quintic equation could be solved by using an elliptic modular function.

R. B. King, *Beyond the quartic equation*. Birkhäuser Boston, Inc., Boston, MA, 1996.

A Glimpse at Galois Theory

Consider the polynomial equation

$$f(t) = t^4 - 4t^2 - 5 = 0.$$

which factorizes as

$$(t^2 + 1)(t^2 - 5) = 0.$$

So there are four roots $t = i, -i, \sqrt{5}, -\sqrt{5}$.

These form two natural pairs: i and $-i$ go together, and so do $\sqrt{5}$ and $-\sqrt{5}$. Indeed, it is impossible to distinguish i from $-i$, or $\sqrt{5}$ from $-\sqrt{5}$, by algebraic means, in the following sense.

Notice that it is possible to write down some polynomial equations with rational coefficients that is satisfied by some selection from the four roots.

For example, if we let

$$\alpha = i \quad \beta = -1 \quad \gamma = \sqrt{5} \quad \delta = -\sqrt{5}$$

then such equations include

$$\alpha^2 + 1 = 0, \quad \alpha + \beta = 0, \quad \delta^2 - 5 = 0, \quad \gamma + \delta = 0, \quad \alpha\gamma - \beta\delta = 0$$

and so on.

There are infinitely many valid equations of this kind. On the other hand, infinitely many other algebraic equations, such as $\alpha + \gamma = 0$, are manifestly false.

Experiment suggests that if we take any valid equation connecting α, β, γ, and δ, and interchange α and β, we again get a valid equation.

The same is true if we interchange γ and δ. For example, the above equations lead by this process to

$$\beta^2 + 1 = 0 \quad \beta + \alpha = 0 \quad \gamma^2 - 5 = 0 \quad \delta + \gamma = 0$$

$$\beta\gamma - \alpha\delta = 0 \quad \alpha\delta - \beta\gamma = 0 \quad \beta\delta - \alpha\gamma = 0$$

and all of these are valid.

In contrast, if we interchange α and γ, we obtain equations such as

$$\gamma^2 + 10 \quad \gamma + \beta = 0 \quad \gamma + \beta = 0$$

which are false.

The operations that we are using here are *permutations* of the zeros $\alpha, \beta, \gamma, \delta$. In the usual permutation notation, the interchange of α and β is

$$R = \begin{pmatrix} \alpha & \beta & \gamma & \delta \\ \beta & \alpha & \gamma & \delta \end{pmatrix}$$

and that of γ and δ is

$$S = \begin{pmatrix} \alpha & \beta & \gamma & \delta \\ \alpha & \beta & \delta & \gamma \end{pmatrix} \; .$$

If these two permutations turn valid equations into valid equations, then so must the permutation obtained by performing them both in turn, which is

$$T = \begin{pmatrix} \alpha & \beta & \gamma & \delta \\ \beta & \alpha & \delta & \gamma \end{pmatrix} \; .$$

Are there any other permutations that preserve all the valid equation? Yes, or course, the identity

$$I = \begin{pmatrix} \alpha & \beta & \gamma & \delta \\ \alpha & \beta & \gamma & \delta \end{pmatrix} \ .$$

It can be checked that only these four permutations preserve valid equations; the other 20 all turn some valid equation into a false one.

- These four permutations form a **group**, which we denote by G.

- What Galois realized is that the structure of this group to some extent controls how we should set about solving the equation.

Consider any quartic polynomial $g(t)$ with the **same** Galois group G and denote its zeros again by $\alpha, \beta, \gamma, \delta$.

Consider three subfields of $\mathbb{C}$ related to $\alpha, \beta, \gamma, \delta$, namely,

$$\mathbb{Q} \subseteq \mathbb{Q}(\gamma, \delta) \subseteq \mathbb{Q}(\alpha, \beta, \gamma, \delta)$$

Let $H = \{I, R\} \subseteq G$ which is a subgroup of G. Assume that we also know the following two facts:

1. The numbers fixed by H are precisely those in $\mathbb{Q}(\gamma, \beta)$.

2. The numbers fixed by G are precisely those in $\mathbb{Q}$.

Then we can work out how to solve the quartic equation $g(t) = 0$, as follows.

The numbers $\alpha+\beta$ and $\alpha\beta$ are obviously both fixed by H.

By fact (1), $\alpha+\beta$ and $\alpha\beta$ lie in $\mathbb{Q}(\gamma,\delta)$. But since

$$(t-\alpha)(t-\beta) = t^2 - (\alpha+\beta)t + \alpha\beta$$

this means that α and β satisfy a quadratic equation whose coefficients are in $\mathbb{Q}(\gamma,\delta)$.

That is, we can use the formula for solving a quadratic to express α, β in terms of rational functions of γ and δ, together with nothing worse than square roots. Thus we obtain α and β as radical expressions in γ and δ.

To find γ and δ, notice that $\gamma + \delta$ and $\gamma\delta$ are fixed by the whole of G; they are clearly fixed by R, and also by S, and these generate G.

- Therefore, γ and δ and $\gamma\delta$ belong to $\mathbb{Q}$ by fact (2) above.

- Hence, γ and δ satisfy a quadratic equation over $\mathbb{Q}$, so they are given by radical expressions in rational numbers.

- Plugging these into the formulas for α and γ we find that all four zeros are radical expressions in rational numbers.

We have not found the formulae explicitly, but we have shown that certain information about the Galois group necessarily implies that they exist. This example illustrates that the subgroup structure of the Galois group G is closely related to the possibility of solving the equation $g(t) = 0$.

Galois discovered that this relationship is very deep and detailed.

A polynomial is solvable by radicals if and only if its Galois group is solvable.

H.M. Edwards, *Galois theory.* Graduate Texts in Mathematics, 101. Springer-Verlag, New York, 1984. J.P. Tignol, *Galois' theory of algebraic equations.* World Scientific Publishing Co., Inc., River Edge, NJ, 2001.

J. Rotman, *Galois theory*. Second edition. Universitext. Springer-Verlag, New York, 1998.

I. Stewart, *Galois Theory*. Third edition. Chapman & Hall/CRC Mathematics. Chapman & Hall/CRC, Boca Raton, FL, 2004.

M. Kuga, *Galois' dream: group theory and differential equations.* Birkhäuser Boston, 1993.

What do we mean by solving a polynomial equation ?

Compare the equations

$$x^2 = 5 \quad \text{and} \quad x^2 = -1.$$

• i is a solution of the latter is simply by definition.

• For $\sqrt{5}$, it is non-trivial that there is a real number that squares to 5.

• When we say we can "solve" the equation $x^2 = 5$, we may also mean we are able to prove that a unique positive real solution exists and $\sqrt{5}$ is just the name that we give to this solution.

What do we mean by solving a polynomial equation ?

Meaning III:

We can show that some roots or zeros exist in certain given number field.

• In this sense, we can solve all polynomial equations within the field of complex numbers.

• This is the so-called Fundamental Theorem of Algebra (FTA) which says that

every non-constant complex polynomial has at least one complex zero.

• The existence of $real$ roots of an equation of odd degree with $real$ coefficients is quite clear.

Since a real polynomial of odd degree tends to oppositely signed infinities as the independent variable ranges from one infinity to the other.

It follows by the connectivity of the graph of the polynomial that the polynomial must assume a zero at some point.

• In general, it is not clear, for example, why at least one solution of the equation

$$x^3 = 2 + \sqrt{-121}$$

is of the form $a + bi, a, b \in \mathbb{R}$.

This problem was considered by the Italian mathematician Bombelli in 1560 when he tried to solve the equation

$$x^3 - 15x = 4$$

which has a real solution 4.

Indeed, by applying the cubic formula, he obtained

$$x \quad = \quad \sqrt[3]{2+\sqrt{-121}} - \sqrt[3]{-2+\sqrt{-121}}.$$

He then proposed a “wild” idea that

$$\sqrt[3]{2+\sqrt{-121}} = 2 + b\sqrt{-1},$$

where b remains to be determined.

Cubing both sides, he showed that $b = 1$.

Similarly, he found out that $\sqrt[3]{-2+\sqrt{-121}} = 2 - \sqrt{-1}$ so that

$$x = 2 + \sqrt{-1} - (-2 + \sqrt{-1}) = 4.$$

- Many books assert that the complex numbers arose in mathematics to solve the equation $x^2 + 1 = 0$, which is simply not true. In fact, they originally arose as in the above example.

- Another impetus towards the Fundamental Theorem of Algebra came from calculus.

Since complex roots to real polynomial equations came in conjugate pairs, it was believed by the middle of the seventeenth century that

every real polynomial could be factored over the reals into linear or quadratic factors.

It was this fact that enabled the integration of rational functions by factoring the denominator and using the method of partial fractions.

Johann Bernoulli asserted in a 1702 paper that such a factoring was always possible, and therefore all rational functions could be integrated.

Interestingly, in 1702 Leibniz questioned the possibility of such factorizations and proposed the following counter-example:

$$\begin{aligned} x^4+a^4 &= (x^2+a^2\sqrt{-1})(x^2-a^2\sqrt{-1}) \\ &= \left(x+a\sqrt{\sqrt{-1}}\right)\left(x-a\sqrt{\sqrt{-1}}\right)\left(x+a\sqrt{-\sqrt{-1}}\right)\left(x-a\sqrt{-\sqrt{-1}}\right). \end{aligned}$$

Leibniz believed that since no nontrivial combination of the four factors yielded a real divisor of the original polynomial, there was no way of factoring it into real quadratic factors.

He did not realize that these factors could be combined to yield $x^4+a^2 = (x^2-\sqrt{2}ax+a^2)(x^2\sqrt{2}ax+a^2)$. It was pointed out by Niklaus Bernoulli in 1719 (three years after the death of Leibniz) that this last factorization was a consequence of the identity $x^4+a^4 = (x^2+a^2)^2-2a^2x^2$.

It is well known that Albert Girard stated a version of the Fundamental Theorem of Algebra in 1629 and that Descartes stated essentially the same result a few years later.

Attempts to Prove the FTA:

i) Jean le Rond d'Alembert (1746, 1754)

ii) Leonhard Euler (1749)

iii) Daviet de Foncenex (1759)

iv) Joseph Louis Lagrange (1772)

v) Pierre Simon Laplace (1795)

vi) James Wood (1798)

vi) Carl Friedrich Gauss (1799, 1814/15, 1816, 1848)

Gauss in his Helmstedt dissertation gave the first generally accepted proof of FTA.

C.F. Gauss, ”Demonstratio nova theorematis functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi poss” (A new proof of the theorem that every rational algebraic function in one variable can be resolved into real factors of the first or second degree), Dissertation, Helmstedt (1799); Werke 3, 130 (1866).

Gauss (1777-1855) considered the FTA so important that he gave four proofs.

i) 1799 (discovered in October 1797), a geometric/topological proof.

ii) 1814/15, an algebraic proof.

iii) 1816, used what we today know as the *Cauchy integral theorem*.

iv) 1849, used the same idea in the first proof.

In the introduction of the fourth proof, Gauss wrote "the first proof] $\cdots$ had a double purpose, first to show that all the proofs previously attempted of this most important theorem of the theory of algebraic equations are unsatisfactory and illusory, and secondly to give a newly constructed rigorous proof." (English translation by D.E. Smith, *Source book in mathematics*, McGraw-Hill, New York,pp.292-293)

• The proofs of d'Alembert, Euler, and Foncenex all make implicit use of the FTA or invalid assumptions.

• All the pre-Gaussian proofs of the FTA assumed the existence of the zeros and attempted to show that the zeros were complex numbers.

• Gauss's was the first to avoid this assumption of the existence of the zeros, hence his proof is considered as the first rigorous proof of the FTA.

• However, according to Stephen Smale (Bull. Amer. Math. Soc. **4** (1981), no. 1, 1–36), Gauss's first proof assumed a subtle topological fact and there actually contained an immense gap and even though Gauss redid this proof 50 years later, the gap remained. It was not until 1920 that Gauss's proof was completed by A. Ostrowski.

- Moreover, it is also now possible to repair d'Alembert and Lagrange's proofs, see for example,

C. Baltus, D'Alembert's proof of the fundamental theorem of algebra. *Historia Math.* **31** (2004), no. 4, 414–428

J. Suzuki, Lagrange's proof of the fundamental theorem of algebra. *Amer. Math. Monthly* **113** (2006), no. 8, 705–714.

- Nowadays, there are many different proofs of the FTA, see for example,

B. Fine and G. Rosenberger, *The fundamental theorem of algebra*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1997.

• Five main approaches to prove the FTA.

i) topological (the winding number of a curve in $\mathbb{R}^2$ around 0).

ii) analytic (Liouville's theorem: bounded entire function must be constant).

iii) algebraic (every odd degree polynomial with real coeff. has a real zero).

iv) probabilistic (results on Brownian motions).

v) nonstandard analysis.

M.N. Pascu, A probabilistic proof of the fundamental theorem of algebra. *Proc. Amer. Math. Soc.* **133** (2005), no. 6, 1707–1711

G. Leibman, A nonstandard proof of the fundamental theorem of algebra. *Amer. Math. Monthly* **112** (2005), no. 8, 705–712.

From FTC to FTA

We shall prove the FTA by applying the Fundamental Theorem of Calculus (FTC):

Let $f : [a, b] \to \mathbb{R}$ be continuous then

A) There exists some function $F : [a, b] \to \mathbb{R}$ such that

$$\frac{dF}{dx}(x) = f(x).$$

B) If $\frac{dF}{dx}(x) = f(x)$, then

$$\int_a^b f(x)dx = F(b) - F(a).$$

By Liouville's theory of integration in finite terms, we know that $\int e^{-t^2} dt$, the anti-derivative of the function e^{-t^2} cannot be expressed "explicitly" (or "in closed form") in terms of "elementary functions" which are built up by using the variable and constants, together with repeated algebraic operations and the taking of exponentials and logarithms.

On the other hand, from part A of the FTC, we know that the anti-derivative of e^{-t^2} exists on any finite interval.

Exercise. Take $f(x) = |x - 1|,\ x \in [0, 2]$. By FTC (part A), there exists some function F such that

$$\frac{dF}{dx}(x) = f(x).$$

Can you find this F?

R.B. Burckel, Fubinito (Immediately) Implies FTA, *The American Mathematical Monthly*, **113**, No. 4, 344-347. April, 2006.

Proof of FTA. Assume P is a non-constant complex polynomial such that

$$P(z) \neq 0 \quad \forall z \in \mathbb{C}.$$

Set $f = 1/P$. Continuity of the rational function f at 0 implies that

$$\lim_{r \downarrow 0} f(re^{i\theta}) = f(0) \neq 0 \qquad \text{(uniformly in } \theta \text{ on the real line } \mathbb{R}). \tag{1}$$

The (rational) function f is differentiable with respect to its complex variable z; let prime denote that differentiation. Then the chain rule gives

$$D_\rho f(\rho e^{i\theta}) = e^{i\theta} f'(\rho e^{i\theta}), \qquad D_\theta f(\rho e^{i\theta}) = \rho i e^{i\theta} f'(\rho e^{i\theta}).$$

Therefore

$$D_\rho f(\rho e^{i\theta}) = \frac{1}{\rho i}\, D_\theta f(\rho e^{i\theta}). \tag{2}$$

For $0 < r < R < \infty$, by the FTC,

$$\int_{-\pi}^{\pi} \int_{r}^{R} D_\rho f(\rho e^{i\theta}) d\rho \, d\theta = \int_{-\pi}^{\pi} [f(Re^{i\theta}) - f(re^{i\theta})] \, d\theta \tag{3}$$

and

$$\int_{r}^{R} \int_{-\pi}^{\pi} \frac{1}{\rho i} D_\theta f(\rho e^{i\theta}) d\theta \, d\rho = \int_{-\pi}^{\pi} \frac{1}{\rho i} [f(\rho e^{i\pi}) - f(\rho e^{-i\pi})] \, d\rho = 0. \tag{4}$$

The function of (ρ, θ) that appears in (2) is continuous on the compact rectangle $[r, R] \times [-\pi, \pi]$. Hence, can apply Fubini's theorem to (3) and

(4) and this yields

$$\int_{-\pi}^{\pi} [f(Re^{i\theta}) - f(re^{i\theta})]\, d\theta = 0 \qquad (0 < r < R < +\infty). \tag{5}$$

Since P is a **non-constant polynomial**, $f = 1/P$ would satisfy

$$f(Re^{i\theta}) \to 0 \qquad \text{(uniformly in } \theta \in \mathbb{R} \text{ as } R \to +\infty).$$

In that case, from (1) and (5) with $R = 1/r \to +\infty$ would follow

$$\int_{-\pi}^{\pi} [0 - f(0)] d\theta = 0, \tag{6}$$

contradicting the fact $f(0) \neq 0$. Hence P must have a zero in $\mathbb{C}$. □

What do we mean by solving a polynomial equation ?

With no hope left for the exact solution formulae, one would like to compute or approximate the zeros of polynomials.

Meaning IV: Try to approximate the zeros with high accuracy.

In general, we would like to find some iterative algorithms to the approximate the zero with high accuracy at a low computational cost (use less time and memory).

Newton's Method

Many algorithms have been developed which produce a sequence of better and better approximations to a solution of a general polynomial equation. In the most satisfactory case, iteration of a single map, Newton's Method.

Newton's method was first defined in Newton's *De methodis serierum et fluxionum* (written in 1671 but first published in 1736).

Newton's map: Let p be a non-linear polynomial with degree n, the Newton's map of p is defined as

$$N_p(z) = z - \frac{p(z)}{p'(z)}.$$

It is known that if we choose an initial point z_0 in $\mathbb{C}$ *suitably* and let

$$z_{n+1} = N_p(z_n) = z_n - \frac{p(z_n)}{p'(z_n)}, n = 0, 1, ...,$$

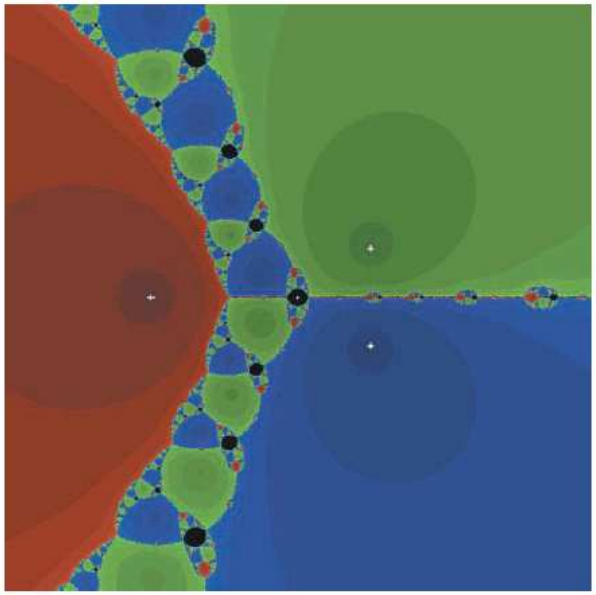
then the sequence $\{z_n\}$ will converge to a zero of p which is also a fixed point of N_p.

• For degree two polynomials, Schröder and Cayley independently proved that there was a line separating the two roots such that any initial guess in the same connected component of a root converges to that root.

• Thus, Newton's Method, converges to a zero for almost all quadratic polynomials and initial points; it is a "generally convergent algorithm."

• But for degree 3 polynomials it converges too infrequently.

- For example, consider the cubic polynomial $p(z) = z^3 - 2z + 2$.



The above figure shows the Newton map N_p over the complex numbers. Colors indicate to which of the three roots a given starting point converges; black indicates starting points which converge to no root, but to the superattracting 2-cycle $(0 \to 1 \to 0)$ instead.

With examples like this, Stephen Smale raised the question as to whether there exists for each degree a generally convergent algorithm which succeeds for all polynomial equations of that degree.

Curtis T. McMullen answered this question in his PhD thesis (1985), under Dennis Sullivan, where he showed that no such algorithm exists for polynomials of degree greater than 3, and for polynomials of degree 3 he produces a new algorithm which does converge to a solution for almost all polynomials and initial points.

One can obtain radicals by Newton's method applied to the polynomial

$$f(x) = x^d - a,$$

starting from any initial point.

In this way, solution by radicals can be seen as a special case of solution by generally convergent algorithms.

This fact led Doyle and McMullen to extend Galois Theory for finding zeros of polynomials. This extension uses McMullen's thesis together with the composition of generally convergent algorithms (a “tower”).

They showed that the zeros of a polynomial could be found by a tower if and only if its Galois group is nearly solvable, extending the notion of solvable Galois group.

• As a consequence, for polynomials of degree bigger than 5 no tower will succeed. While for degree 5, Doyle and McMullen were able to construct such a tower.

J. Shurman, *Geometry of the quintic.* John Wiley & Sons, 1997.

Since McMullen has shown that there are no generally convergent purely iterative algorithms for solving polynomials of degrees 4 or greater, it follows that there is a set of positive measure of polynomials for which a set of positive measure of initial guesses will not converge to any root with any algorithm analogous to Newton's method.

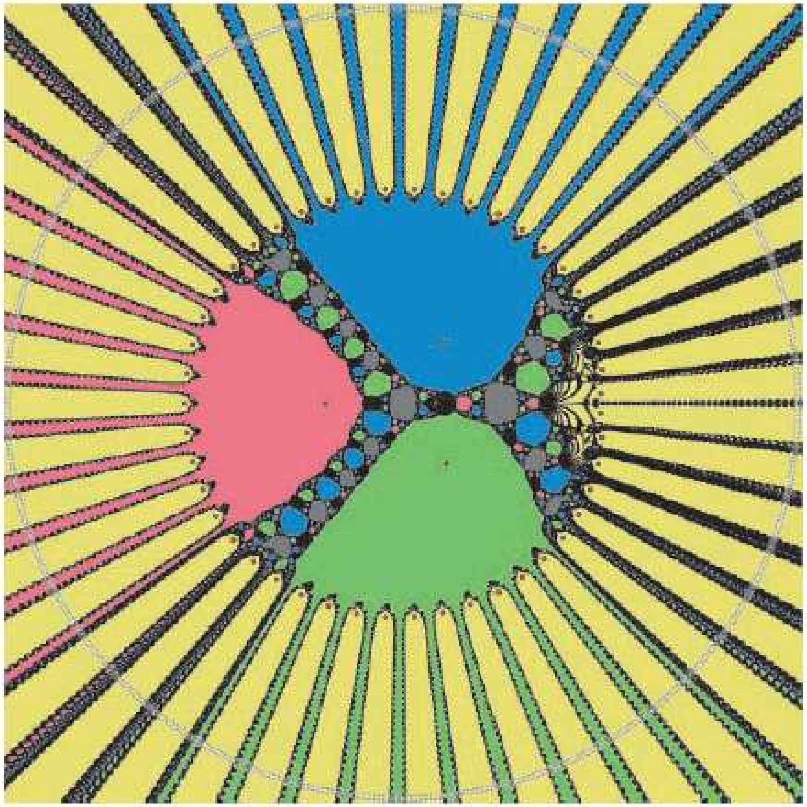
On the other hand, the following important paper shows how to save the Newton's method.

J. Hubbard, D. Schleicher, S. Sutherland, How to find all roots of complex polynomials by Newton's method. *Invent. Math.* **146** (2001), no. 1, 1–33.

The authors proved that, for each degree d, there exists a universal set S_d of initial guesses such that, for any polynomial of degree d (suitably normalized) and any of its roots, at least one point in S_d converges to it.

This set has approximately $1.11 d \log^2 d$ points which are equally distributed on a small fraction of $\log d$ circles and the set S_d can be constructed explicitly.

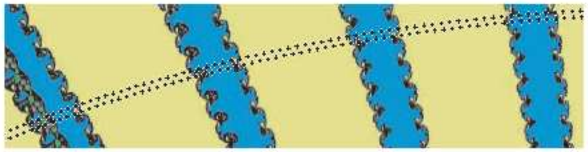
S_{50} for $p(z) = z^{50} + 8z^5 - \frac{80}{3}z^4 + 20z^3 - 2z + 1$



For the degree 50 polynomial, $z^{50}+8z^5-\frac{80}{3}z^4+20z^3-2z+1$, a set of starting points S_{50} as specified by the previous result, is indicated by small crosses distributed on two large circles.

There are 47 zeros near the unit circle, and 3 zeros well inside, all marked by red disks. There is also an attracting periodic orbit (basin in grey).

A close-up is shown below.



Stephen Smale, 1930- :

An important result in Mathematics is never finished.

Richard Hamming, 1915-1998:

Mathematics is nothing but clear thinking.

Thank You